Guest lectures will be recorded.october 11 and october 13.

Cross\_entropy

Y log (p)

Y = predictive values

P = probs

Forward pass: given theta (init guess) find error

Backward pass: update weights according to SGD

Habitat Paper:

How long does it take to complete one iteration of Forward and Backward pass

Non linear function does not have a convex 🡪 not closed form solution

Numerical methods to minimize: *Conjugate gradenad* and *Gradient descent*

Minibatch = you evaluate the gradient on a part of the dataset.

These days, people uses big batches (as much as your memory can take)

Bigger batch, better representation

Forward and backward per each batch.